Extending THREDDS middleware to serve OGC community

S. Nativi, B. Domenico, J. Caron, E. Davis, L. Bigagli

Adv. Geosci., 8, 57–62, 2006
www.adv-geosci.net/8/57/2006/

Advances in Geosciences

Extending THREDDS middleware to serve OGC community

S. Nativi[1,2], **B. Domenico**[3], **J. Caron**[3], **E. Davis**[3], **and L. Bigagli**[1,2]

[1]Institute of Methodologies for Environmental Analysis of the National Research Council (IMAA-CNR), Italy
[2]Faculty of Engineering, University of Florence at Prato, Italy
[3]Unidata Program Center University Corporation for Atmospheric Research (UCAR), USA

Received: 15 September 2005 – Revised: 30 January 2006 – Accepted: 15 February 2006 – Published: 6 June 2006

Abstract. As far as interoperability is concerned, in a service-oriented framework, it is possible to distinguish different service tiers; each tier contains systems and tools which implement that tier's specific task. The present paper analyses such a framework for the Earth Sciences and the GIS information communities. For the Earth Sciences community, the heterogeneity of existing protocols and data models is outlined, considering the experience of the Unidata community. For the GIS community, the interoperability opportunities laid by the OGC's specifications are briefly introduced. The need of achieving the two communities' frameworks interoperability, and its importance for science Digital Library applications are introduced. A solution is presented and discussed; it is based on the following technologies: THREDDS Data Server, OGC WCS/WFS and ncML-GML. An OGC interoperability experiment, which tests the proposed solution, is briefly presented.

1 Introduction

In the Unidata community framework of client/server data and metadata access systems, there are a number of client/server protocols in use at different data provider sites. At the other end, some client applications can access data via some of the protocols while others can only access data via other protocols. THREDDS catalogs provide information about which datasets are available via which services/protocols (Domenico et al., 2002). The three main client/server (as opposed to full-file transfer with FTP or GridFTP) protocols for remote data access in use in the community are OPeNDAP (Open-source Project for a Network Data Access Protocol), ADDE (Abstract Data Distribution Environment), and netCDF access via HTTP protocol. In many cases the data access systems are augmented and integrated with THREDDS (Thematic Real-time Environmental Distributed Data Services) catalog services which proved inventory list and metadata access. Thus client applications can learn what's available on the site via the THREDDS interface, then access the datasets themselves via OPeNDAP, ADDE, or netCDF/HTTP protocols.

The Unidata netCDF data model (Unidata) is very popular among scientists in the Earth sciences community. The Open Geospatial Consortium's (OGC) Web Coverage Service (WCS) (Evans, 2003) can be used to dispatch netCDF datasets via standard protocols to client applications used by a variety of user groups: the scientific digital library community, the GIS community, as well as the broader Earth science research and education community. All these potential user groups benefit by having netCDF as one form of coverage supported by WCS. Among the main reasons for bringing the two together are:

– The netCDF is one of the most commonly used format for storing the output of weather and climate forecast models. The output of these models is different from many of the other datasets currently supported by in the GIS community. These datasets represent several parameters (e.g., temperature, pressure, wind speed and direction) that vary in three spatial dimensions and involve two distinct time scales (model run time and forecast times).

– The netCDF interface is evolving in a direction that aims to support access to many different file formats (e.g. HDF5, GRIB, GINI, McIDAS AREA, NEXRAD, netCDF-3, netCDF-4, etc.) via several different client/server protocols (e.g., OPeNDAP, ADDE and HTTP) that are already established in the atmospheric and ocean sciences data provider community.

Correspondence to: S. Nativi
(nativi@imaa.cnr.it)

Published by Copernicus GmbH on behalf of the European Geosciences Union.

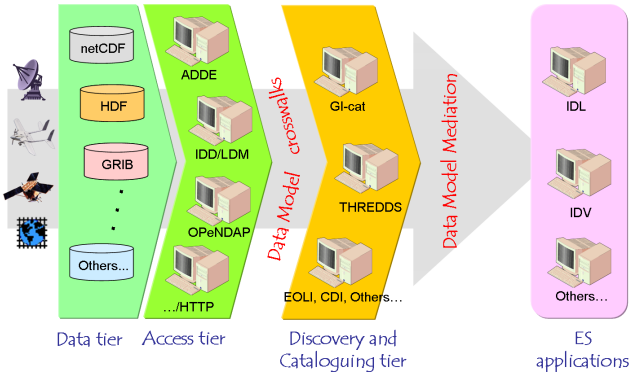

Fig. 1. Service-oriented interoperability framework for the ES information community.

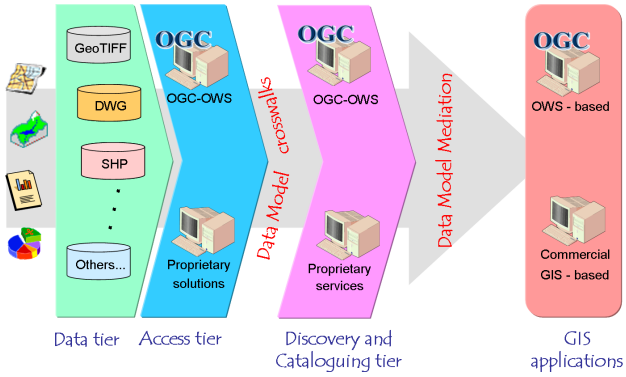

Fig. 2. Service-oriented interoperability framework for the GIS information community.

– OGC's WCS can be successfully used to dispatch netCDF datasets via standardized protocols to many interested groups:
  – the Earth Science information Community (many of whom now use netCDF interfaces);
  – the GIS Information Community;
  – the general public;

Thus the inclusion of netCDF as a WCS format adds only one new data access interface that in turn provides access to collections of forecast model output via a variety of protocols that are already in use in the data provider community. In a more general perspective, there is need for well-accepted binary encoding of WCS Coverages; specifying WCS profiles (e.g. the netCDF profile of WCS).

The following paragraphs present and discuss a service-oriented view which accommodates the heterogeneity of Unidata community client/server data and metadata access systems. The introduced architecture aims at achieving interoperability between earth sciences and GIS communities, leveraging netCDF and WCS specifications.

2 Interoperability and service-oriented architectures

In a service-oriented (SO) framework, as far as interoperability is concerned, it is possible to distinguish among a generic application tier for consumers and three main service tiers for service providers: 1) data tier; 2) access tier; 3) discovery and cataloguing tier.

In a given community, each tier contains systems and tools which implement that tier's specific task, such as: data managing services, data access services, data discovery and cataloguing services, etc. Fig. 1 depicts such a framework for the Earth Sciences (ES) community.

In this framework, interoperability must be implemented between service tiers, sorting out the heterogeneity of systems and tools which characterize each task. Heterogeneity exists in the service data model and interface and binding technology. Hence, data model crosswalk and/or mediation solutions must be conceived, implemented, and tested in order to achieve effective interoperability. The essential role of these middleware solutions is outlined in Figs. 1 and 2.

2.1 The Earth Sciences information community

As far as ES information community is concerned, the SO framework is characterized by a truly heterogeneous situation for different systems and tools at different tiers. The Data tier is characterized by the presence of services working with heterogeneous data models, such as: HDF, netCDF, GRIB, GINI, McIDAS AREA, NEXRAD, etc. The Access tier consists of numerous and well-accepted services, using different models and interfaces, such as: ADDE, IDD/LDM, OPeNDAP, etc. The Discovery and Cataloguing tier is characterized by a relatively small number of services; most of them are still under refinement, such as: THREDDS catalog services, GI-cat (Bigagli et al., 2004), etc. Lastly, consumer applications can be very heterogeneous, such as: an IDL-based application, a MATLAB-based procedure, UNIDATA IDV, etc.

The Data and Access tiers are generally tightly-coupled, using data model mediation/crosswalk solutions which are well-accepted and used and can be considered part of services. In order to make Access and Discovery and Cataloguing fully interoperable, it is necessary to develop specific middleware paying particular attention to data model crosswalks. A crosswalk is used for translating between different models (e.g. data models, metadata formats, etc.); in a mediation solution, a neutral third party (the mediator) performs this task making use of a more general and common model. Indeed, existing access and cataloguing services are generally characterized by different data model, especially as far as semantics content is concerned. Data model mediation middleware is necessary in order to implement fully interoperability between consumer applications and the Discovery and Cataloguing tier services.

2.2 The GIS information community

As far as the GIS community is concerned, the SO framework can be simplified distinguishing between proprietary solutions and OWS (OGC Web Services)-based solutions (Whiteside, 2005). Data tier is characterized by heterogeneous data models, such as: SHP, DWG, GeoTIFF, etc. For both Access and Discovery and Cataloguing tiers, it is possible to divide services in two categories: OGC-OWS and proprietary solutions, respectively. The same simplification can be used for describing consumer application services.

Data model mediation and crosswalk middleware is necessary for implementing interoperability between Access and Discovery and Cataloguing services, as well as Discovery and Cataloguing and consumer applications tiers. As a matter of fact, OGC specifies most of the middleware useful to loosely couple Access and Discovery and Cataloguing services (e.g. GML). Figure 2 depicts the interoperability framework for the GIS information community.

3 An interoperability framework for: ES, GIS and DL info communities

ES and GIS information realms present several common aspects and share common data: they both generate, manage and use geospatial data. Nevertheless, as it was previously pointed out, they have been developing their own semantics, schemas and tools, which very often are not effectively interoperable (Nativi et al., 2004). The science Digital Library (DL) presently must access ES & GIS through separate solutions, but there is need for effective use of these related geo-information and geospatial data together. Therefore, there exists the need for a common interoperability framework.

The present work introduces a solution for implementing an interoperability framework which aims at fulfilling the following main objectives:

- To respond to ES and DL community requests for desktop client access to data on multiple remote servers.
- To implement standards-based web services approach to accessing and providing gridded data, and its metadata.
- To foster more interdisciplinary data uses.
- To integrate GIS clients and servers for providing access to societal impacts, infrastructure, hydrology data, etc.

In order to achieve such infrastructure, it is necessary to address the important open issue previously discussed: implementation of standards-based data models and service protocols crosswalk and mediation solutions. Specifically, introducing interoperability solutions between existing tiers – which gather heterogeneous services- complying with some essential requirements:

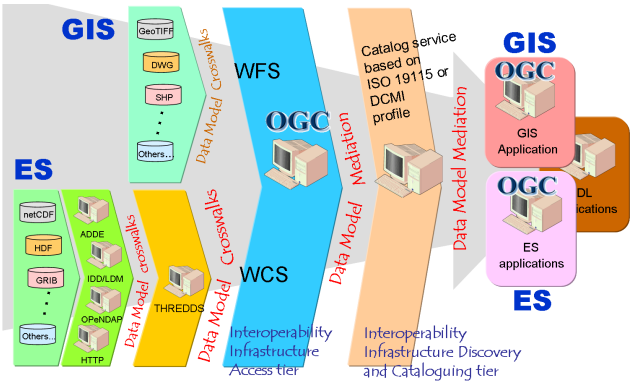

Fig. 3. The proposed solution as common service-oriented interoperability framework.

- Allow components de-coupling, enabling the framework extensibility.
- Build on existing heterogeneous and well-accepted services.
- Facilitate service chaining and composability.

3.1 The proposed interoperability architecture

Figure 3 depicts the introduced approach. Referring to Figs. 1, 2 and 3, it is apparent that the proposed architecture considers the following specific solution: use the GIS Access tier to access datasets from both GIS and ES realms. It becomes the Access tier for the entire interoperability infrastructure. OGC OWS interfaces are adopted: WCS for ES datasets and WFS (Vretanos, 2005) for GIS ones, respectively. For the GIS realm, the common Access tier is easily interfaced with the Data tier, using consolidated middleware. On the other hand, for the ES realm, the common infrastructure Access tier cannot be directly interfaced to the Data tier. In fact, WCS needs to interoperate with high level services, whose data model captures and formalizes aggregated aspects of ES dataset structure and semantics. Services grouped in the ES Discovery and Cataloging tier could provide such high level interface and data model. The proposed solution considers the services offered by THREDDS.

The proposed infrastructure mainly addresses GIS and ES interoperability; nevertheless, DL applications will certainly take a great advantage of such an interoperability framework.

For completeness and clarity the Interoperability Infrastructure Discovery Tier is also depicted. It contains any catalog service based on either ISO 19115 or DCMI (Dublin Core Metadata Initiative) profiles (e.g. the OGC CS-W). This tier will not be further detailed in the following paragraphs, which focus on the middleware components introduced by this infrastructure for making ES realm tiers interoperable.

Summarizing, the proposed solution is based on the following main technological choices:

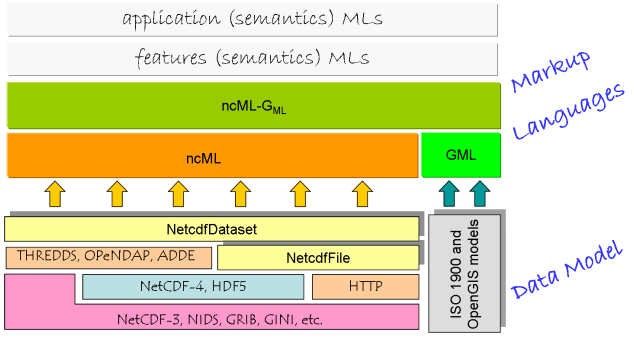

Fig. 4. NetCDF data model and the used XML information set.

– THREDDS data server along with netCDF interface;

– OGC WCS/WFS and GML.

As explained in the following paragraphs, a third technology plays an important role to make the proposed solution work: ncML−G_{ML}. According to its name, it is a piece of middleware which interfaces netCDF and GML data spaces.

4 Tiered interoperability implementation

According to Fig. 3, it is noteworthy that a couple of data model crosswalks and protocol mediation middleware are essential to enable the introduced architecture:

1. Middleware to enable interoperability between Discovery-Cataloguing and Access tiers for ES realm;

2. Middleware to enable interoperability between GIS Access and ES Discovery-Cataloguing tiers

4.1 Interoperability between Catalog and Access/Data tiers for ES realm

The first mediation middleware is implemented by the THREDDS Data Server (TDS) (Caron[a]). TDS uses the netCDF implementation of the CDM interface, working out access protocols and data model mediations.

TDS is an augmented web server that provides metadata and data access for scientific datasets, building on and extending a number of existing technologies:

1. THREDDS Dataset Inventory Catalog (Domenico et al., 2002) used to provide virtual directories of available data and their associated metadata.

2. The Netcdf-Java library (Caron[b]) to read NetCDF, OpenDAP, and HDF5 datasets, as well as other binary formats such as GRIB and NEXRAD into a "Common Data Model" (CDM)[1]. Figure 4 shows the netCDF implementation of the CDM.

3. An integrated OPeNDAP server. OPenDAP (OPenDAP) is a widely used, subsetting data access method built on the HTTP (web) protocol.

4.2 Interoperability between GIS Access and ES Catalog-discovery tiers

The second mediation middleware is implemented by a specific component called: THREDDS-WCS gateway (Domenico[a]). It implements a THREDDS – WCS Data Model crosswalk, as well as a THREDDS – WCS Protocols mediation. On this topic, there exists an ongoing OGC IE (Interoperability Experiment), called GALEON (Geo-interface for Atmosphere, Land, Earth, and Ocean netCDF) (OGC).

In order to achieve the required interoperability, a specific implementation of THREDDS-WCS gateway is required, which implements a data model mediation component based on the ncML-GML model (Nativi, 2005b; Domenico et al., 2005). This mediates between THREDDS and GIS-oriented data models, such as OGC's WCS or WFS.

4.2.1 NcML−G_{ML}

NcML−G_{ML} (hereafter referred as NcML-GML) is an extension of the netCDF Markup Language (ncML) (Nativi, 2005a; Caron et al., 2005); it enables encoding of netCDF datasets in terms of GML elements. It was conceived to facilitate interoperability between the fluid earth sciences and the GIS communities. NcML-GML leverages the ncML ability to encode multi-dimensional arrays, and the wide acceptance of GML for encoding geo-spatial metadata. Presently, the ncML-GML version 0.6.2 is based on GML 3.1 grammar; it implements one of the possible encodings of a general interoperability model (i.e. both abstract and content interoperability models) which reconciles typical atmospheric sciences and geo-spatial data models.

NcML-GML is particularly useful where both ES and GIS semantics and metadata content are required. That's the case of services/applications which aim at serving both communities. A valuable example could be the OGC's WCS and WFS. As a matter of fact, the GALEON IE is intended to demonstrate such a capability (Domenico[b]).

NcML-GML's main objectives are:

[1]CDM is an abstract data model that the netCDF (Unidata), HDF5 (NCSA) and OpenDAP (University of Rhode Island) developers are using to converge their respective data models towards. The CDM also adds "Georeferencing Coordinate Systems" and specialized "Scientific Data Type" layers, which provides the semantics needed to convert datasets to other protocols and formats such as those required by GIS systems.

1. to explicitly mediate from netCDF hyperspatial data model to geo-information coverage data model (i.e. a mediation ML and content model reconciliation schema for ES and GIS info realms);

2. to explicitly encode netCDF CF conventions into GML-based elements (i.e. an extension of ncML core schema, based on GML grammar).

Figures 4 and 5 depict the ncML/ncML-GML data model and the adopted mediation solution, respectively. As depicted in Fig. 5, for the ES information domain, we consider the netCDF-CF data model, enriched by the THREDDS metadata elements and encoded using the ncML model. On the other hand, for the GIS information domain, we consider the WCS protocol data model which is a GML application which, in turn, is the encoding model for the ISO 19100 data model. In this framework, the ncML-GML acts as the mediation language to implement the crosswalk between the netCDF-CF and ISO 19100 data models (i.e. the coverage model).

A complete description of the mediation tasks implemented by ncML-GML is reported in (Nativi et al., 2005).

4.2.2 GALEON

The OGC "Geo-interface for Atmosphere, Land, Earth, and Ocean netCDF" (GALEON) Interoperability Experiment supports open access to the output of atmospheric and oceanographic modeling and simulations (OGC). The GALEON IE will implement a geo-interface to netCDF datasets via the OpenGIS Web Coverage Server (WCS 1.0) protocol specification. The interface will provide interoperability among netCDF, OPeNDAP, ADDE, and THREDDS client/server and catalog protocols. The IE may generate change requests to the WCS and other OpenGIS specifications.

The GALEON IE will implement a geo-interface to netCDF datasets via the WCS 1.0 protocol specification. It will implement the WCS as a layer above a set of client/server and catalog protocols already in use in the atmospheric and oceanographic sciences communities. In particular, it will leverage OPeNDAP servers that provide access to netCDF datasets and accompanying THREDDS servers providing ancillary information about the datasets. The IE will investigate the feasibility of adapting data and metadata originating from OPeNDAP/THREDDS servers to the WCS specifications, in so contributing to bridge the gap between the atmospheric, oceanographic and GIS communities, by alleviating data interoperability issues. This experiment can be seen as a step in the direction of interoperability with data systems already in existence in the oceanographic and atmospheric sciences.

The initiators of GALEON are: Unidata/UCAR, IMAA-CNR, George Mason University, and the NASA – Geospatial Interoperability Office.

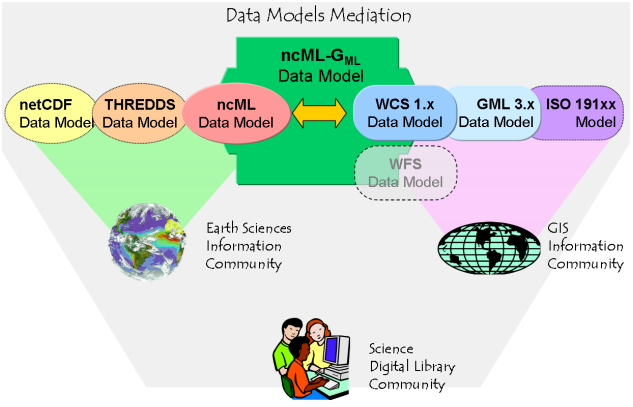

Fig. 5. The adopted data model mediation approach.

Three use cases have been considered, according to the following interaction steps:

– THREDDS inventory catalogs are accessed and used to generate dataset lists for the getCapabilities response

– NcML-GML metadata from the THREDDS server are used to construct the describeCoverage response.

– NetCDF objects are retrieved from the netCDF/OPeNDAP server either directly via the netCDF interface or from an OPeNDAP service.

– The netCDF objects are then transformed into one of the following three forms for transmission back to the client in response to the get coverage request:

  1. geoTIFF binary file;
  2. ncML-GML document (containing a netCDF/OPeNDAP pointer, a file pointer, a SOAP attachment or data itself);
  3. netCDF binary file.

5 Conclusions

ES services framework is characterized by a high level of heterogeneity. For the GIS community it is possible to distinguish between OGC and not-OGC services. ES and GIS information realms present common aspects and share common data; nevertheless, they have been developing their own semantics, schemas and tools. DL presently must access ES & GIS through separate solutions, but there is need for effective use of these related geo-information and geospatial data together.. Therefore, there exists the need for a common interoperability framework.

The present works introduces a possible interoperability infrastructure. The proposed architecture considers the following specific solution: to use the GIS Access tier, adopting only OGC OWS interfaces (i.e. WCS/WFS), to access

datasets from both GIS and ES realms. A couple of data model crosswalks and protocol mediation middleware are essential to enable the introduced architecture:

1. Middleware to enable interoperability between Discovery-Cataloguing and Access tiers for ES realm;
2. Middleware to enable interoperability between GIS Access and ES Discovery-Cataloguing tiers

Considered middleware components are based on the following main technological choices:

– THREDDS data server along with netCDF interface;

– OGC WCS/WFS and GML;

– ncML-GML.

This solution mainly addresses GIS and ES interoperability; nevertheless, DL applications will certainly take a great advantage of such an interoperability framework.

Interoperability between OGC and THREDDS middleware is extremely important for ES, GIS and DL communities. The effectiveness of this objective is going to be assessed by the OGC's GALEON Interoperability Experiment.

Edited by: E. Cutrim, M. Ramamurthy, S. Nativi and L. Miller
Reviewed by: anonymous referees

References

Bigagli, L., Nativi, S., Mazzetti, P., and Villoresi, G.: GI-Cat: a Web Service for Dataset Cataloguing Based on ISO 19115", Proceedings of the 15th International Workshop on Database and Expert abase and Expert Systems Systems Applications, Zaragoza (Spain), Sep 2004, IEEE Computer Society Press, ISBN 0-7695-2195-9, pp. 846–850.

Caron[a], J. (Ed.): THREDDS Data Server – Status, http://my.unidata.ucar.edu/content/projects/THREDDS/tech/server/ServerStatus.html (last access: 28 April 2006).

Caron[b], J. (Ed.): NetCDF Java Library (Version 2), http://www.unidata.ucar.edu/software/netcdf-java/ (last access: 28 April 2006).

Domenico, B., Caron, J., Davis, E., Kambic, R., and Nativi, S.: Thematic Real-time Environmental Distributed Data Services (THREDDS): Incorporating Interactive Analysis Tools into NSDL, J. Dig. Inf., 2, 4, Art. No. 114, 29 May 2002.

Domenico[a], B.: WCS-THREDDS Gateway, http://www.unidata.ucar.edu/projects/THREDDS/OGC/WCS-THREDDS\%20Gateway.htm (last access: 28 April 2006).

Domenico[b], B. (Ed.): GALEON (Geo-interface to Atmosphere, Land, Earth, Ocean netCDF) homepage, http://galeon-wcs.jot.com/WikiHome (last access: 28 April 2006).

Evans, J. (Ed.): OpenGIS® Web Coverage Service (WCS) Implementation Specification, OpenGIS Implementation Specification, OGC document number 03-065r6, October 2003.

Nativi, S., Blumenthal, B., Habermann, T., Hertzmann, D., Raskin, R., Caron, J., Domenico, B., Ho, Y., and Weber, J.: Differences among the data models used by the Geographic Information Systems and Atmospheric Science communities, Proceedings of American Meteorological Society – 20th Interactive Image Processing Systems Conference, Seattle (WA), Jan 2004.

Nativi, S., Caron, J., Davis, E., and Domenico, B.: Design and implementation of netCDF Markup Language (NcML) and Its GML-based extension (NcML-GML), Computers & Geosciences Journal, Elsevier Publication, 31, 9, 1104–1118, November 2005a.

Nativi, S., Domenico, B., Caron, J., Davis, E., and Bigagli, L.: An interoperability language to connect netCDF and Geographic communities: ncML-GML v. 0.5, presented at the GML and Geo-Spatial Web Services Conference 2005, Vancouver, British Columbia, July 2005, submitted, http://my.unidata.ucar.edu/content/projects/THREDDS/GALEON/ncML-GML\%200.5_submitted.pdf, 2005b.

OGC: GALEON IE (GALEON IE), http://www.opengeospatial.org/initiatives/?iid=173 (last access: 28 April 2006).

OPeNDAP Inc.: OPeNDAP homepage, http://www.opendap.org/ (last access: April 2006).

Unidata netCDF Group: netCDF Usage, Unidata software: netCDF, http://my.unidata.ucar.edu/content/software/netcdf/usage.html (last access: 28 April 2006).

Vretanos, P. (Ed.): OpenGIS® Web Feature Service (WFS) Implementation Specification, OpenGIS Implementation Specification, OGC document number 04-094, May 2005.

Whiteside, A. (Ed.): OpenGIS® Web Service Common Implementation Specification, OpenGIS Implementation Specification, OGC document number 05-008c1, May 2005.